



# Il sistema gestionale dei prodotti della ricerca

**M. D'Alessandro - L. Sanelli - M. Gattari  
A. Paoletti - F. Serafini**

Direzione Sistemi Informativi - Amm. Centrale INFN  
*In collaborazione con GLV e GLOS INFN*

DOI: [10.15161/oar.it/143237](https://doi.org/10.15161/oar.it/143237) - GenOA week - ottobre 2023



# Agenda



Introduzione e contesto



Concept



Focus su Machine Learning



*What Next*

# Introduzione

- Progetto condiviso da DSI, GLV e GLOS per lo sviluppo di un sistema gestionale dei **prodotti** della ricerca INFN
- **Obiettivo:** sviluppare un sistema di raccolta e organizzazione di (meta)dati e metriche ad uso delle Commissioni Scientifiche e del Gruppo di Valutazione INFN.
- **Adeguamento tecnologico e funzionale** di un preesistente archivio attivo dai primi anni 2000





# Contesto

- **2004**: Sviluppo interno di un archivio di pubblicazioni INFN per la **valutazione** dell'attività scientifica
- Supporto ad esercizi annuali e pluriennali di analisi dati (VTR, VQR...) e valutazione della **produttività** scientifica degli **esperimenti**
- Sistema *fortemente tagliato sulle esigenze dell'Ente* ma sviluppato con tecnologie, UI/UX e paradigmi superati (limitanti)
- **2023**: Nuova **VQR**, nuove specifiche, nuova visione, nuovo sistema

# Concept



Nuova UI/UX



Data model flessibile



Un «indice» degli autori



Machine Learning

# UI/UX: nuova interfaccia

PHP V7

Pagina 1 / 160 - Trovati 3193 record

Le pubblicazioni asse

INFN Prodotti della Ricerca

Antonello Paoletti

Ricerca Inserisci Autori Admin

Pubblicazione

Tesi

Talk

Colonne (10/13): Autori, File, Rivista, Sigle, Autori %...

	Autori	Titolo	Rivista	Tipo	Anno	File	Autori %	Sigla/CSN	Etichette	Fonti	Azioni
<input type="checkbox"/>	Christensen, RB et al.	AN ALGORITHMIC APPROACH TO ENTANGLEMENT-ASSISTED QUANTUM ERR... WOS:000742873200001   10.3934/amc.2021072									
<input type="checkbox"/>	Adam, J et al.	Surface Interactions Studies of Novel Two-Dimensional Molybd... WOS:000793922000001   10.1080/00032719.2022.2070186									
<input type="checkbox"/>	Amendola, A et al.	On shear motions in nonlinear transverse isotropic elastodyn... WOS:000823600400001   10.1177/10812865221105452									
<input type="checkbox"/>	Burgarth, D et al.	Control of Quantum Noise: On the Role of Dilations WOS:000830372200001   10.1007/s00023-022-01211-y									
<input type="checkbox"/>	Torresi, L et al.	Shockwave and spallation in silver and other materials by su... WOS:000852447600001   10.1002/ctpp.202200092									
<input type="checkbox"/>	Murari, A et al.	Information theoretic and neural computational tools for met... WOS:000854734900001   10.1007/s00521-022-07768-3									
<input type="checkbox"/>	Evans, C et al.	The CUBES science case WOS:000862545100001   10.1007/s10686-022-09864-7									
<input type="checkbox"/>	Barrillon, P et al.	The EUSO@TurLab project in the framework of the JEM-EUSO pro... WOS:000864985300001   10.1007/s10686-022-09871-8									
<input type="checkbox"/>	Mazzucchi, S et al.	Chernoff approximations of Feller semigroups in Riemannian m... WOS:000866533000001   10.1002/mana.202100291									
<input type="checkbox"/>	Pedrielli, A et al.	In search of the ground-state crystal structure of Ta2O5 fro... WOS:000868424000002   10.1016/j.commtsci.2022.111828									
<input type="checkbox"/>	Acampora, G et al.	D-NISQ: A reference model for Distributed Noisy Intermediate... WOS:000868900700002   10.1016/j.inffus.2022.08.003									
<input type="checkbox"/>	Guardiani, C. et al.	Computational methods and theory for ion channel research	ADVANCES IN PHYSICS-X	Review	2022			DYNSYSMATH/4		WOS, SCOPUS	Elimina >>
<input type="checkbox"/>	Bolognesi, Stefano et al.	Anomalies and phases of strongly coupled chiral gauge theories: Recent developments	INT J MOD PHYS A	Review	2022		100	common4/4		WOS	Elimina >>
<input type="checkbox"/>	Bini, Donato et al.	Radiation-reaction and angular momentum loss at the second post-Minkowskian order	PHYS REV D	Article	2022		50	QGSKY/4		WOS	Elimina >>
<input type="checkbox"/>	Paradisi, Paride et al.	High-energy frontier of the muon g-2 at a muon collider	PHYS REV D	Article	2022		67	common4/4		WOS	Elimina >>
<input type="checkbox"/>	Hu, M. et al.	Ultrafast laser surface irradiation of silicon: Effects of repetition rate in vacuum and air	APPL SURF SCI	Article	2022		38	common5/5		WOS	Elimina >>
<input type="checkbox"/>	Agostino, Rocco et al.	Forecasting constraints on deviations from general relativity in f (Q) gravity with standard sirens	PHYS REV D	Article	2022		50	QGSKY/4		WOS	Elimina >>

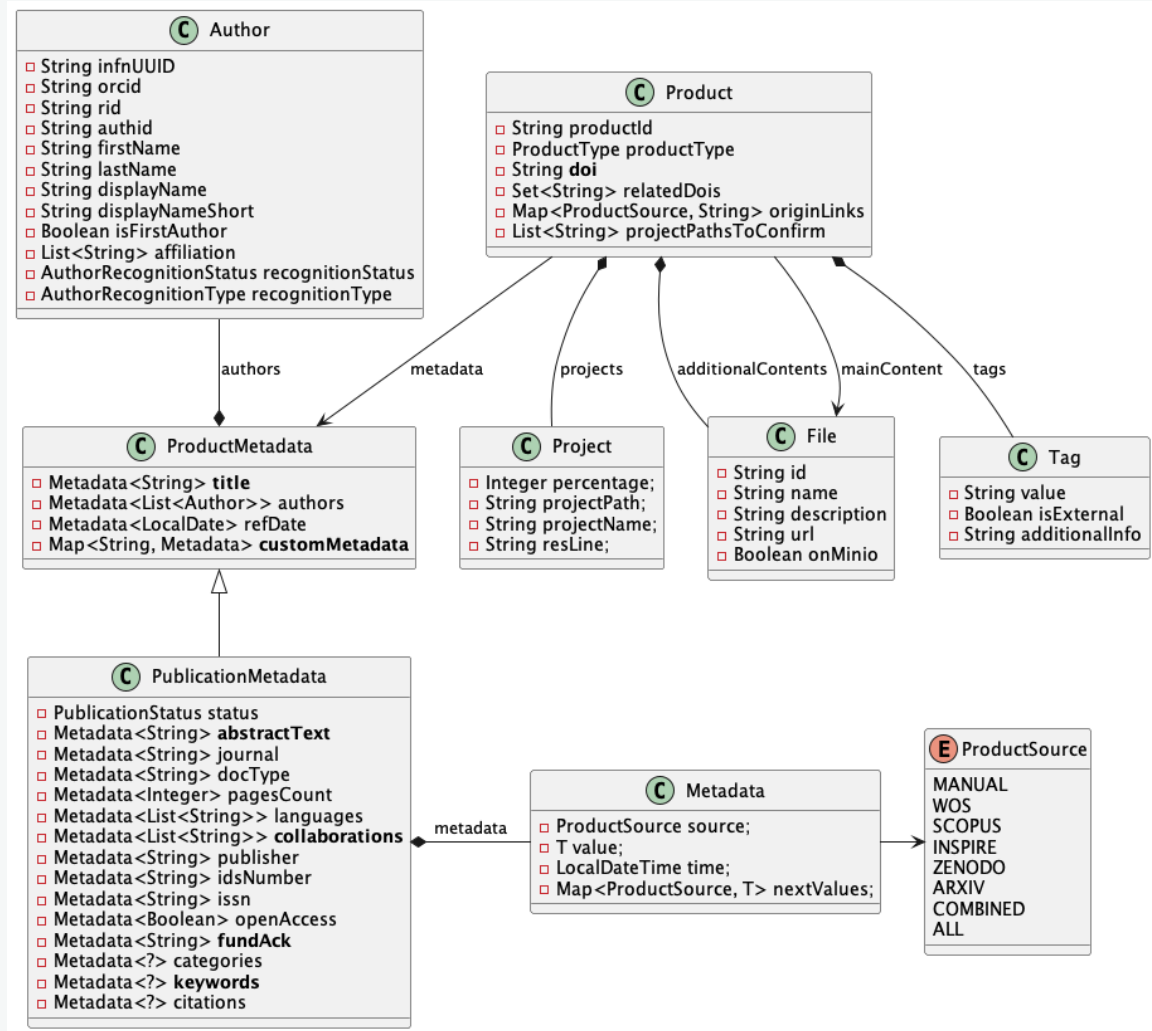


# Data Model

- Passaggio da un modello dati relazionale a struttura fissa a una struttura variabile **non relazionale**
- Popolamento del modello dati a partire da **fonti** dati eterogenee (ISI vs Inspire) con politiche di **conversione, priorità e coesistenza** dei singoli metadati definite a sistema.
- Salvataggio dei record originali e possibile (ri) definizione delle politiche di **convergenza** dei metadati (anche dopo l'aggiunta di nuove fonti)
- Metadati aggiuntivi ed etichette



# Prodotto della ricerca







# Gestione degli autori



Popolamento di un «indice» con informazioni sugli autori a seguito di nuovi articoli/autori importati nel sistema (con conseguente riconoscimento delle *stringhe* autori)



Acquisizione dei principali **identificativi** esterni ed interni all'Ente



Memorizzazione delle diciture inerenti **l'affiliazione** e il **display name** per un dato autore

Utile per velocizzare e supportare i successivi processi di riconoscimento

# UI/UX: Gestione Autori

Dettaglio Autori infn

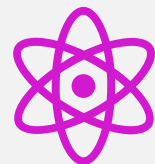
Nominativo	Email	Affiliazioni	INFN UUID - ORCID	Azione
Antonelli, M.	-	INFN, Frascati, Italy Lab Nazi Frascati, Frascati, Italy	0000-0002-2293-5726	associa
Calvetti, M.	-	INFN Sez Pisa, Pisa, Italy Univ Pisa, Dipartimento Fis E Fermi, Pisa, Italy	0000-0003-0125-2165	associa
Cavalli, N.	-	Univ Bologna, Dipartimento Fis & Astron A Righi, Bologna, It.. INFN Sez Bologna, Bologna, Italy		
Cerrito, L.	-	INFN Sez Roma Tor Vergata, Rome, Italy Univ Roma Tor Vergata, Dipartimento Fis, Rome, Italy		
Chomont, A. R.	-	INFN Sez Roma, Rome, Italy Sapienza Univ Roma, Dipartimento Fis, Rome, Italy		
Coimbra, A. E. C.	-	INFN Sez Milano, Milan, Italy Univ Milan, Dipartimento Fis, Milan, Italy		
D'Amico, V.	-	INFN Sez Roma Tre, Rome, Italy Univ Roma Tre, Dipartimento Matemat & Fis, Rome, Italy		
De Santis, M.	-	INFN Sez Roma Tor Vergata, Rome, Italy Univ Roma Tor Vergata, Dipartimento Fis, Rome, Italy		
Fauci Giannelli, M.	-	INFN Sez Roma Tor Vergata, Rome, Italy Univ Roma Tor Vergata, Dipartimento Fis, Rome, Italy		

Riconoscimento Autore

Cerca un'anagrafica corrispondente a **Antonelli, M.**

Nome	Cognome	Struttura	
Cristian	Antonelli		Riconosci
Emma	Antonelli		Riconosci
Gianni	Antonelli		Riconosci
Mara	Antonelli		Riconosci
Marco	Antonelli		Riconosci
Mario	Antonelli	LNF	Riconosci
Matias	Antonelli	TS	Riconosci

# AI/ML



Aggancio progetti di ricerca



Alfa: Riconoscimento autori

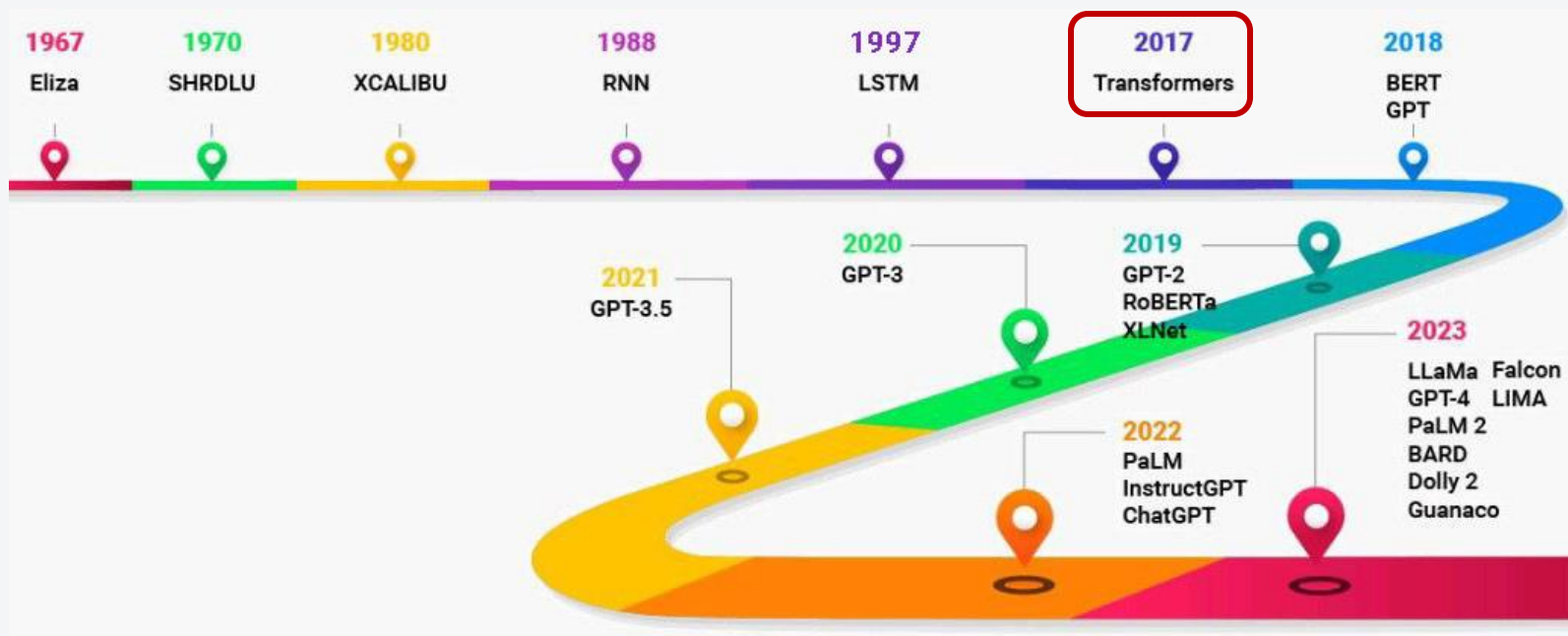


R&D: Aggancio VOR/AAM

# AI/ML

## NLP

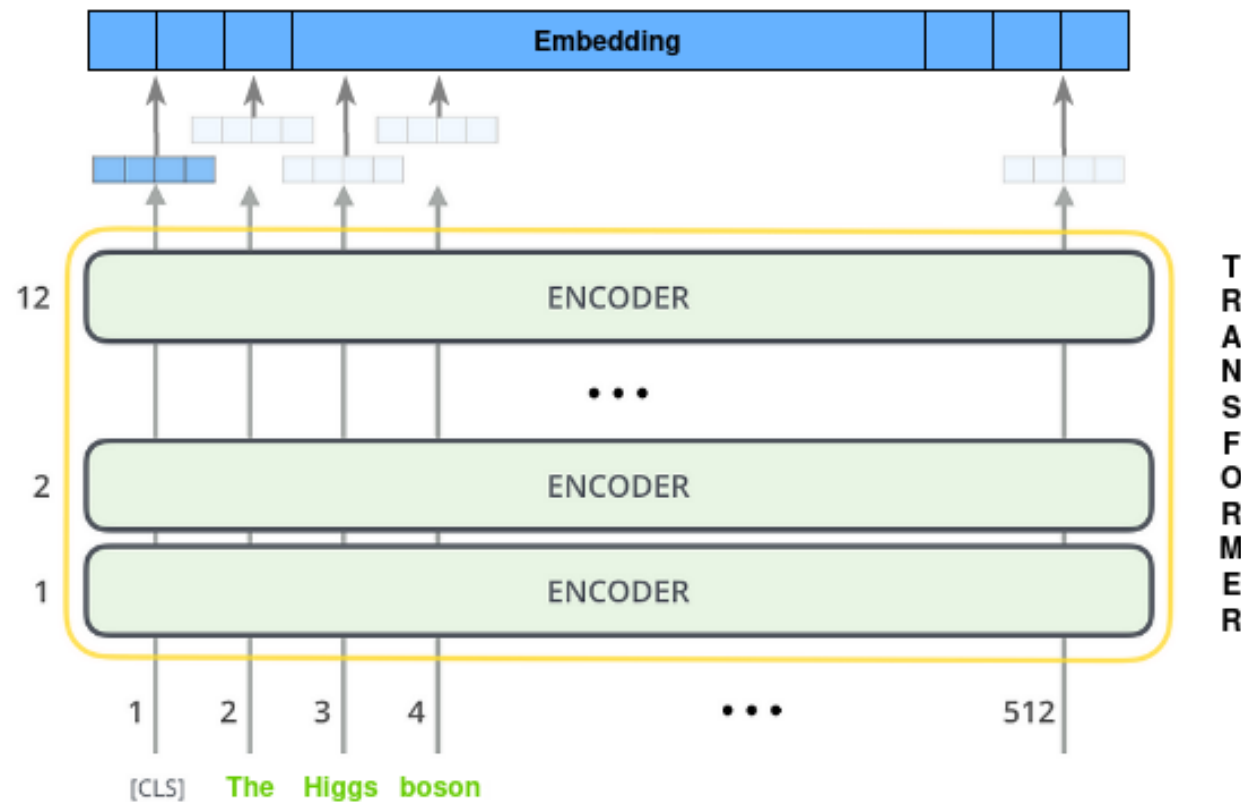
- **NLP** (Natural Language Processing): campo della linguistica e del ML dedicato alla comprensione del linguaggio umano.
- Un evento chiave nella storia dei Language Models è l'introduzione dell'architettura **Transformer** nel 2017 (Google Brain team).
- Transformers si sono dimostrati particolarmente efficaci nel risolvere problemi NLP:
  - **ChatGPT** => Generative Pretrained Transformer



# Transformer

## Architettura

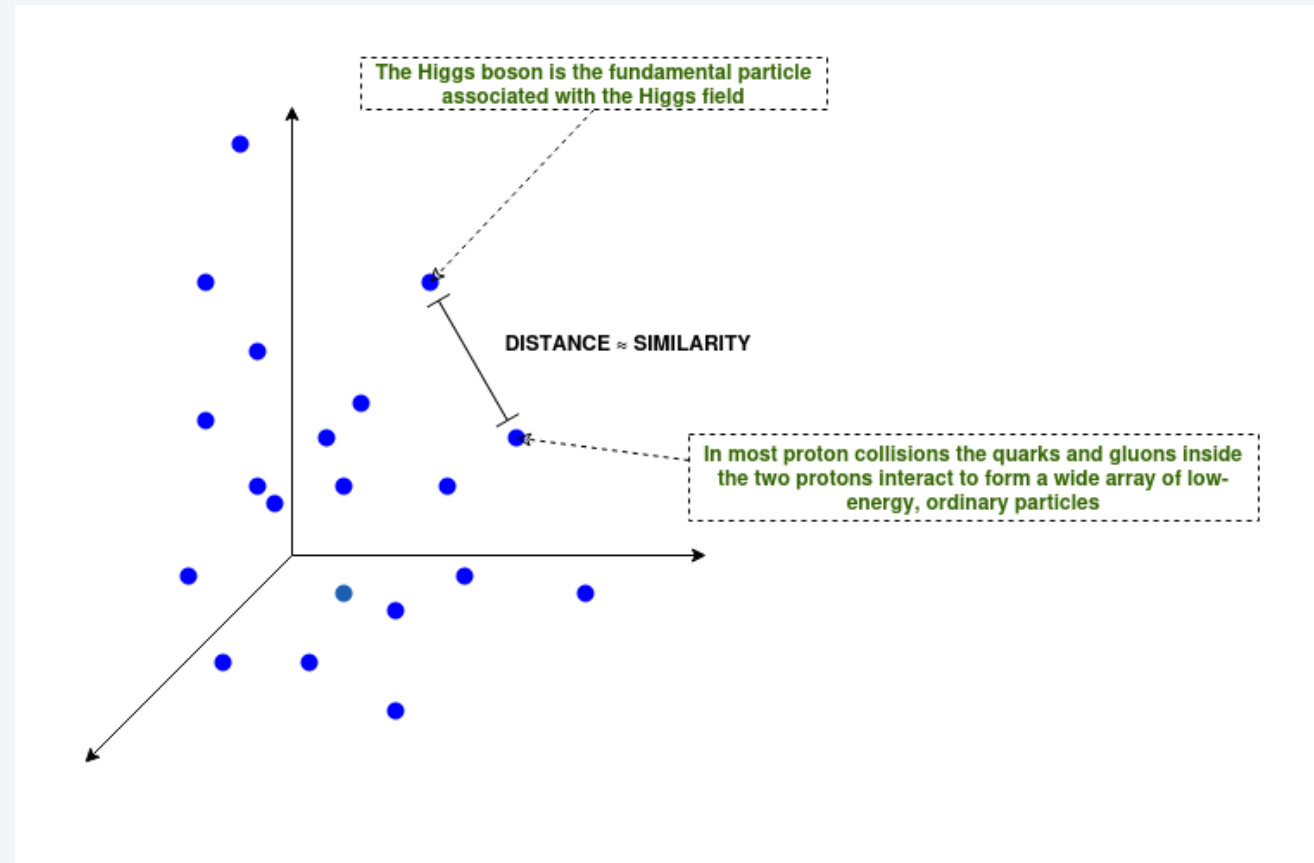
- Transformer: trasforma un testo di input in oggetto matematico, detto **embedding**
- Un embedding è un **vettore** n-dimensionale che rappresenta matematicamente la semantica della frase di input
- Se opportunamente addestrato, vettori "**vicini**" corrispondono a testi semanticamente "**similari**"



# Transformer

## Caso d'uso: riconoscimento sigle di progetto

- **Training** del Transformer su dati di dominio («Linguaggio della fisica»):
  - 60K abstracts articoli INFN
  - 3 giorni training (Nvidia Tesla T4)
- Creazione spazio vettoriale formato dagli embeddings prodotti dal Transformer: vettori "**vicini**" corrispondono ad articoli semanticamente "**similari**"
- Ricerca semantica nello spazio vettoriale per aggancio di nuovi articoli





# What next



Supportare i prossimi esercizi di **valutazione** interni ed esterni (VQR 2020-2024)



Potenziare l'integrazione con servizi interni ed esterni all'INFN per favorire lo **scambio** e l'**analisi** dati



Ampliare i campi di applicazione degli algoritmi di **ML** nei processi di *data enrichment*

# Grazie per l'attenzione

*antonello.paoletti  
marzio.dalessandro  
luca.sanelli  
mauro.gattari  
francesco.serafini  
@infn.it*

